

NeuralMesh™ Unifies Datasets, Accelerates Genomics Research

NeuralMesh™ brings all sequencing data into a single namespace to remove storage silos, feeds servers in parallel to maximize GPU and CPU utilization, and reduces cost per computation for genomics workloads.

Challenges

- Billions of files per directory delay concurrent genomics research jobs
- Sequencing data scatters across disparate data repositories
- Idle GPUs raise research costs
- Retention mandates drive endless capacity growth
- Consortium sites wait on data copies

Solution

NeuralMesh™ distributes every file and its metadata across all nodes, so one read pulls from many drives at once. Clients move I/O directly to GPU and CPU servers, with no staging to local SSD. Distributed metadata handles billions of files in one directory and trillions cluster-wide. S3 and POSIX run on the same physical data blocks, not a gateway, so every pipeline stage shares one copy. Data reduction extends usable capacity without slowing writes. Tiering spans flash and object storage with no manual tuning.

Benefits

- Run more concurrent research jobs
- Consolidate disparate file systems into a single namespace
- Drive GPU utilization past 90%
- Stretch usable capacity up to 6x with guaranteed data reduction
- Federate data across hybrid cloud with on-demand hydration

Spend less money on storage and more time on genomics research

Genomics pipelines move from sequencer to variant call across petabytes of raw reads. When job queues back up, that path stretches from days into weeks, and the delay hurts scientific breakthroughs. A rare disease diagnosis waits. A cohort gets re-queued behind another re-run. A grant-funded GPU cluster sits idle. **Geneticists end up managing infrastructure instead of doing science.** That is not how breakthroughs happen.

NeuralMesh removes data bottlenecks. Metadata is shared across every node, so a directory holding billions of files behaves like any other directory. Physical and logical tenant isolation gives each lab its own slice of the cluster, so one re-run does not stall others.

Among WEKA customers, the **Oklahoma Medical Research Foundation (OMRF)** leverages NeuralMesh to run GATK-based NGS pipelines on GPU-enabled servers. With NeuralMesh, OMRF **cut research job times by 10x**, with some jobs **dropping from 70 days to just a week** and others from weeks to hours. And all of these research workloads no longer required staging data to local SSDs before a run.

“With WEKA, we have so much performance and expandable capacity that we don’t have to think about it.”

— Stuart Glenn, Assistant Director, Oklahoma Medical Research Foundation

